DOCUMENT RESUME

ED 465 283 FL 027 303

AUTHOR	Chang, Yuh-Fang
TITLE	EFL Teachers' Responses to L2 Writing.
PUB DATE	2002-00-00
NOTE	21p.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	College Faculty; *English (Second Language); Foreign Countries; Grading; Higher Education; *Language Teachers; Native Language Instruction; Rating Scales; Scores; Second Language Instruction; Writing (Composition); *Writing Evaluation; *Writing Skills
IDENTIFIERS	Chinese Speaking; Taiwan

ABSTRACT

This study investigated differences in the product and process of evaluating second language compositions by Taiwanese speakers of English. It examined whether such factors as language background (native English speaker versus native Chinese speaker), academic discipline, and educational background affected raters' scoring outcomes; whether rating scales (holistic versus analytic) affected raters' scoring outcomes; and whether raters' holistic scales correlated with specific features of analytic scores. Researchers selected a composition written by a Taiwanese student in a freshmen composition course. A group of 4 native English-speaking and 10 native-Chinese-speaking faculty members read and rated the composition using 2 holistic and 2 analytic grading scales and corrected everything that appeared ungrammatical. The think-aloud process was used to examine the rating process. Results found no significant differences in the score results of the four rating scales between raters of different academic disciplines or educational backgrounds. The mean score of the two groups of raters was significantly different on the Test of Written English rating scale. Significant correlations were found between holistic and analytic scores for content and organization. The mean scores of different rating scales differed significantly. Raters differed in total number of comments and number of factors commented upon. (Contains 9 tables and 18 references.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 465 283

Title: EFL teachers' responses to L2 writing

Author: Yuh-Fang Chang, assistant professor at Providence University

Address: Providence University

English Department

200 Chung-Chi Rd. Sha-lu

Taichung County, Taiwan

Phone: (O) 886-4-2632-8001 ext. 2213

(H) 886-4-2461-2425

E-mail: yfchang2@pu.edu.tw

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Yuh-Fang Chang

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

FL027303

Abstract

When it comes to evaluating composition, one of the major concerns for researchers and test administrators has been inter-rater and intra-rater reliability because the grading behavior of raters varies. Several studies dealing with raters' grading behavior have found that factors such as age, academic discipline, and L1 background affect subjects' responses to writing errors. Besides variation from raters, another key factor complicating the issue of grading behavior is the rating scale (holistic or analytic) adopted to evaluate the composition. Although a considerable body of literature exists addressing these issues, most studies have examined raters in ESL contexts, while relatively little has been done to explore raters in EFL contexts other than Japan. In addition, the focus of most studies has been on the product of assessment. The rating process has received much less attention. The purpose of this study is to investigate the degree to which differences exist in both the product and process of L2 composition evaluation by raters in an EFL setting--Taiwan.

Introduction

Evaluating L2 composition is a time-consuming yet essential task for writing instructors. It is time-consuming because L2 composition often requires feedback not only in content but also in language use. It is essential because writing teacher's feedback (corrective or evaluative) plays a significant role in students' learning achievement. When it comes to evaluating composition, one of the major concerns for researchers and test administrators has been inter-rater and intra-rater reliability because the grading behavior of raters varies. Several studies dealing with raters' grading behavior have found that factors such as age, academic discipline, and L1 background affect subjects' responses to writing errors (Brown, 1991; Freeman, 1981; Janopoulos, 1992; Kobayashi, 1992; Santo, 1988; Song and Caruso, 1996; Vann, Meyer and Lorenz, 1984). Besides variation from raters, another key factor complicating the issue of grading behavior is the rating scale (holistic or analytic) adopted to evaluate the composition (Charney, 1984; Grobe, 1981; Harris, 1977; Homburg, 1984; Nold and Freeman, 1977; Stewart and Grobe, 1979). Although a considerable body of literature exists addressing these issues, most studies have examined raters in ESL contexts, while relatively little has been done to explore raters in EFL contexts other than Japan. In addition, the focus of most studies has been on the product of assessment. The rating process has received much less attention.

The purpose of this study is to investigate the degree to which differences exist in both the product and process of L2 composition evaluation by raters in an EFL

setting--Taiwan. In Taiwan, English is taught as the principal foreign language. It is a required subject from the seventh grade up. As to the product of L2 composition evaluation, the following questions were posed:

(1) Do factors such as L1 background (English native speakers vs. Chinese native speakers), academic disciplines (linguistics, literature, TESOL), and education background (master degree vs. doctoral degree) affect raters' scoring outcomes? (2) Do rating scales (holistic vs. analytic) affect raters' scoring outcomes? 3) Are raters' holistic scores correlated with certain features of the analytic scores?

METHODS

Materials

One composition was selected from among approximately 70 Taiwanese students who were taking a freshman composition course taught by the researcher last semester and who had written their mid-term essay in a formal test environment with a 40-minute time limit (see Appendix A). The selection of the composition used in the study was made on the basis of the following criteria. First, the composition chosen scored in the middle range, representing the writing proficiency level of the majority of students taking the freshman composition taught by the researcher. The second consideration was that the composition contained errors, such as subject-verb agreement and run-on sentences, commonly made by English learners in Taiwan.

The composition selected consisted of 17 sentences (including fragments), with a total of 210 words. To eliminate the possibility that handwriting might affect the

raters' grading behavior, the original composition was kept unmodified but typed double-spaced by the student writer herself.

Subjects

A total of 14 full-time faculty members of the English Department at one university in Taiwan, four native speakers of English and ten native speakers of Chinese, participated in the study (see Table 1). They ranged in age from 32 to 53. Among the four English native speakers, two held degrees in linguistics, one in literature and one in TESOL. Of the ten non-native speakers of English, three held degree in linguistics, three in literature and four in TESOL. There were 13 females and 1 male. They were selected because of their availability at the time of data collection and their willingness to participate in the study.

The exclusion of participants from other institutions in Taiwan was intended to prevent the differences in teachers' expectations from different institutions becoming a confounding variable in the study, because all universities in Taiwan are hierarchically ranked. Those who teach at the English Department of a top level university are likely to differ from those who teach at a university ranked at the middle or bottom ranges in their expectations of a freshman's composition. This difference may, in turn, result in differences in grading outcomes.

Insert Table 1 about here

Procedures

After having sought the participation of the 14 faculty members, the researcher scheduled to meet with each subject to collect the data separately. In order to examine whether scoring systems affect raters' evaluation, each subject was provided with four grading scales -- two holistic scales and two analytic scales (see appendix B). They were asked to read and rate the composition both holistically and analytically. The explanation of each scoring system was provided before they started to read the composition.

The two holistic scales employed in this study included 1) the 100-point scale, a scale commonly used in grading writing assignments in Taiwan and 2) the 6-point scale developed by the Educational Testing Service for its Test of Written English (TWE). The two analytic scales used in the study consisted of 1) an ESL composition profile and 2) a sample analytic scale introduced in Reid (1993). The ESL composition profile, one of the most widely used analytic scales, is composed of five weighted components-- content (30 points), organization (20 points), vocabulary (20 points), language use (25 points), and mechanics (5 points). The categorized features of the second analytic scale are 1) introduction (10 points), 2) support (30 points), 3) organization (20 points), 4) style (20 points) and 5) rhetorical stance (20 points).

While they were reading the composition, they were asked to correct everything that seemed ungrammatical or unacceptable to them. To examine the

rating process, the think-aloud protocol was used. During the process of reading, correcting and scoring the composition, the subjects were instructed to verbally comment into a tape recorder. The time each participant spent in grading the paper varied, ranging from 30 minutes to 90 minutes. The think aloud data of each subject were transcribed to qualitatively describe and analyze similarities and differences between raters in the rating process.

RESULTS

Quantitative Results

The first research question asked whether there was a significant difference in the rating outcomes by raters of different L1 backgrounds (NS vs. NNS), education backgrounds (MA vs. PhD) and academic disciplines (linguistics, literature and TESOL). The subjects' ratings of the categorized features in analytic scales were first summed to yield one single mean score. The means and standard deviation of the ratings of four scales were computed and *t*-test was performed to examine the effects of L1 background and education background, while ANOVA was applied to the academic disciplines.

Insert Table 2 about here

Table 2 displays the results of two-tailed *t*-test of the overall mean scores of four rating scales by English native speakers of rater faculty and non-native speakers of rater faculty. An examination of Table 2 shows that although the overall mean

scores by native speakers of English were higher than those by non-native speakers of English on all four rating scales, the difference between two groups in overall mean scores of the four rating scales was not statistically significant except for the holistic TWE rating scale.

Insert Table 3 about here

The means and standard deviations of the four rating scales as scored by faculty holding MA degrees and Ph.D. degrees are presented in Table 3. As can be seen, the MA faculty's scores are lower than those of the Ph.D. faculty's scores on three rating scales (holistic 100, TWE, and 2nd analytic scale). The difference, however, was not statistically significant.

Insert Table 4 and 5 about here

Table 4 shows the means and standard deviations of four rating scales by raters of three academic disciplines. To examine the rating outcomes of the three groups, a one-way ANOVA is shown in Table 5. The results revealed that there was no significant difference in scoring between academic disciplines, although the mean score by the linguistic faculty was the highest among the three rater groups.

The second research question asked whether the faculty's rating outcomes hold consistent no matter which type of rating scale is used. To examine the effect of rating scales, Friedman test was used. In order to be able to examine the difference statistically, the TWE score was transformed into a scale with a total score of 100

because the other scales have a total score of 100. Table 6 presents the mean and standard deviation of the rating outcomes using different rating scales. The results showed that the difference in the mean score of different rating scales was statistically significant. As can be seen in Table 6, the difference between the lowest and the highest score in each scale was more than 20. In the TWE and second analytic scales, the difference among the raters was a difference between pass and fail. Surprisingly, the mean scores of 100-point holistic scale, which does not have any level descriptors and the first analytic scale (ESL composition profile) were quite close, which probably indicates that the presence or absence of a scoring guide does not influence rating outcomes.

Insert Table 6 about here

The third research question asked whether the raters' holistic scores were correlated with certain features of the analytic scale scores given by the same raters. To examine the relationship between the holistic scores and components of analytic scales, the spearman's rho was performed. Table 7 presents the values of spearman correlation coefficient. An asterisk(*) indicates significant correlation at the .05 level. The analyses uncovered a positive correlation between both holistic scores and the analytic scores on features of content and organization in the first analytic scale: ESL composition profile. As to the second analytic scale, the score of the 100-point holistic scale correlated with the score on the features of organization and

mechanics, while the TWE score correlated with the content and organization.

Insert Table 7 about here

Qualitative Results

I. The decision-making process: Raters' Comments

The think-aloud data were transcribed in full by the researcher and grouped into 15 categories, given in Table 8. 15 categories were later combined into three types of comments: language use, contents, and organization. Table 9 shows the total number of comments made by raters, the number of factors commented on and the frequency of types of comments by each rater. Each rater's score was also listed in Table 9. As can be seen, the focus of most participants was on grammatical accuracy. The only exception was a NS rater with a literature degree (rater 4) who had more comments on content than language use. Interestingly, the score she gave was not particularly distinct from the other raters', which suggests that raters may actually give a similar score for different reasons.

While the raters' response pattern resembled each other in that language use was the most frequent commented category by most raters, they differed in the total number of comments, ranging from 11 to 48, and the number of factors commented on, from 5 to 14. When the total number of comments each rater made was compared with each rater's scores, it was found that those who made more comments (or corrections) did not necessarily give lower grades. For instance, rater 1, who

made 36 comments, gave a score of 88, whereas rater 13, who made 13 comments, gave a score of 70. In addition, the raters also differed on the number of factors they commented on. Some raters concentrated on certain categories (e.g., rater 6, 10, 13) while some commented on almost every category (e.g., rater 4, 5, 9). The number of factors they commented on, however, did not seem to play an influential role in the final grade.

Insert Table 8 and 9 about here

II. The decision of the final grade

As mentioned above, the findings revealed that during the process of L2 composition evaluation, the raters differed in the total number of comments and the number of factors they commented on. These, however, did not account for the differences in the final grades given by raters. Further analysis of the think-aloud protocol found that in the final stage of composition evaluation –grading—raters differed in the criteria they applied and the degree of decisiveness with which they assigned a score. When using an analytic scale or a holistic scale with level descriptors (e.g., TWE), raters were identical in that they all read through the description first and then decided the grade. Even though each rater was presented with the same student writing and scales with level descriptors, each rater's perception of how good the student composition was in terms of each component differed. The "mechanics" component of the ESL composition profile (i.e., the first analytic scale)

can illustrate this. The component of mechanics comprises 5 points of the total score. Of the 14 raters, three raters gave 5 points; 9 raters gave 4 points; 1 gave 3.5 points and 1 gave 2 points. Analysis of the think-aloud protocols suggests that the discrepancy among the raters may result from differences in the attention to detail given by each rater and in raters' perception of how serious a certain error is. In the examined student writing, the writer consistently made the mistake of putting a space between the word and punctuation at the end of a sentence. None of the raters who gave 5 points spotted the error. Those who found the error had different opinions on how serious the error was. Most raters considered it to be a minor mistake, while two raters mentioned that it was a terrible mistake because spacing is a basic and fundamental writing convention.

When using the 100-point holistic scale without level descriptors, raters considered the student writer's background in making their final decisions. When considering what they would give out of 100 points, many of them had comments like "If it's high school, it's very good. If it's college, it's passable." or "If a freshman, I probably give sixty or seventy. If this is a junior student, I will give this fifty five and have him rewrite it." One teacher specifically mentioned, "I know a tough grader will give a score of 65; a nice grader, 75. I would like to average and give it 70." Moreover, raters' teaching experience mattered since it determined the criteria teachers used to evaluate the writing. For instance, a novice native speaker rater (1 year EFL experience) stated that "I will give this high 80's probably 88 because this is

much better than the students I have ever taught last semester." His teaching experience was mainly with non-English majors, which probably accounts for his final grade and generous comments about the composition—"good argument, idea well-structured, reasoning generally clear, problems with idiomatic phrasing and word choice, very good overall."

Raters also differed in how decisively they assigned a score, which may result from personality differences. No matter which scale was used, while there seemed to be no problem for raters to locate the score range they intended to give (e.g., fair to poor (21-17 points) in content area), the rating styles differed when it came to assigning a score from within the range they chose. Some of the raters struggled in assigning a specific number (20 or 19) and kept saying, "I don't know" before and after assigning a score, while others gave a score promptly and decisively without too much consideration.

Discussion and Conclusions

This study examines the degree to which differences exist in both the product and process of L2 composition evaluation by raters in an EFL setting--Taiwan. With regard to the product of L2 composition evaluation, unlike previous studies (e.g., Vann, Meyer and Lorenz, 1984; Santos, 1988; Song and Caruso, 1996) which indicate that the academic discipline of faculty members was an important factor affecting raters' responses, the results of the first analysis revealed that there were no significant differences in the score results of the four rating scales between raters of

different academic disciplines or educational backgrounds. In contrast with Connor-Linton's (1995) finding that American ESL raters and Japanese EFL raters gave similar quantitative ratings to the same essay, the mean score of the two groups of teachers (NS raters vs. NNS raters) was significantly different on one holistic scale, the TWE rating scale. In addition, as opposed to Sweedler-Brown's (1993) finding that no correlation existed between holistic scores and analytic scores assigned to organization, the results of this study showed the opposite to be true—significant correlations were found between holistic scores and analytic scores for content and organization. More studies need to be done in order to clarify whether the difference between previous studies and the present study could be attributed to the difference in research settings (EFL vs. ESL).

The results of the second analysis indicate that the rating scale may be an important factor affecting rating outcomes. The finding that the mean scores of different rating scales were significantly different suggests that not every rating scale automatically produces a similar outcome. The difference, as shown in this study, could be between passing and failing. The comparison of two holistic scales revealed that the holistic scale with level descriptors (the TWE scale) did not seem to provide more helpful guidance to the raters than the 100-point holistic scale, since the raters may have different interpretations of the descriptors. With regard to the different rating outcomes of the two analytic scales, the range of the score provided in each component seems to contribute to the difference between the two scales. In the first

analytic scale (ESL composition profile), the lowest score of each component is not 1 (e.g., in content area, the lowest score was 13), which to some extent reduces the possibility of producing extremely different scores between raters. In contrast, the starting point of almost each component in the second analytic scale was 1. In addition, the second analytic scale consists of more components than the first analytic scale, which may magnify the effect of using 1 as a starting point in the rating scale.

With regard to the process of L2 composition evaluation, the qualitative analysis section of this study revealed that the raters differed in the total number of comments and the number of factors they commented on. Hence, the qualitative reasons for the rating outcomes among the raters may not be identical. In addition, unlike Kobayashi's (1992) findings that those who found more errors gave lower ratings than those who failed to find them, the results of the present study showed that those who made more corrections did not necessarily give stricter ratings, which suggests that a student's writing itself may not be the only factor influencing rating outcomes; factors such as differences in raters' expectations of student's writing performance at certain academic levels (e.g., English-major freshman, sophomore etc.) and raters' teaching experience also affect their judgment of the worth of a piece of writing.

Two implications can be drawn from the present study. First, scoring a L2 composition is a ranking procedure, which is an inevitable step in teaching L2 writing. To make the rating outcomes more equitable for students in EFL contexts, it is necessary to provide grading training sessions for all EFL writing instructors in an

English department, especially for novice teachers, to assure every rater shares similar grading philosophy and rating standard, which in turn will increase inter- and intra-rater reliability. An additional implication concerns the rating scales. Rating scales are not equal. Program administrators and L2 writing instructors need to familiarize themselves with the merits and demerits of various types of rating scales and choose one based on their pedagogical and testing needs.

References

Brown, J. (1991). Do English and ESL faculties rate writing samples differently? TESOL Quarterly, 25(4), 587-603.

Charney, D. (1984). The validity of using holistic scoring to evaluate writing: a critical overview. Research in the Teaching of English, 18(1), 65-81.

Chastain, K.(1981). Native speaker evaluation of student composition errors. Modern Language Journal, 65, 288-94.

Connor-Linton, J. (1990). Looking behind the curtain: what do L2 composition ratings really mean? TESOL Quarterly, 24(4), 762-765.

Connor-Linton, J. (1995). Crosscultural comparison of writing standards: American ESL and Japanese EFL. World Englishes, 14(1), 99-115.

Freedman, S.(1981). Influences on evaluators of expository essays: beyond the text. Research in the Teaching of English, 15(3), 245-255.

Grobe, C. (1981). Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. Research in the Teaching of English, 15(1), 75-85.

Harris, W. (1977). Teacher response to student writing: a study of the response patterns of high school English teachers to determine the basis for teacher judgment of student writing. Research in the Teaching of English, 11,175-183.

Homburg, T. (1984). Holistic evaluation of ESL compositions: can it be validated objectively? TESOL Quarterly, 18(1), 87-107.

Janopoulos, M. (1992). University faculty tolerance of NS and NNS writing errors: a comparison. Journal of Second Language Writing, 1(2), 109-121.

Kobayashi, T. (1992). Native and nonnative reactions to ESL compositions. TESOL Quarterly, 26(1), 81-112.

Nold, E. and Freedman, S. (1977). An analysis of readers' responses to essay. Research in the Teaching of English, 11, 164-174.

Rifkin, B. and Roberts, F. (1995). Error gravity: a critical review of research design. Language Learning, 45(3), 511-537.

Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students. TESOL Quarterly, 22(1), 69-90.

Song, B. and Caruso, I. (1996). Do English and ESL faculty differ in evaluating the essays of native English-speaking and ESL students? Journal of Second Language Writing, 5(2), 163-181.

Stewart, M. (1979). Syntactic maturity and mechanics of writing. Research in the Teaching of English, 13, 207-215.

Sweedle-Brown, C.O. (1993). ESL essay evaluation: The influence of sentence-level and rhetorical features. Journal of Second Language Writing, 2(1), 3-17

Vann, R., Meyer, D. and Lorenz, F. (1984). Error gravity: a study of faculty opinion of ESL errors. TESOL Quarterly, 18(3), 427-440.

Table 1. Participants' information

Academic discipline	NS (Ph.D)	NNS (MA.)	NNS (Ph.D)	Total
Linguistics	2	1	2	5
Literature	1	1	2	4
TESOL	1	O	4	5

Table 2. Rating outcomes of NS faculty versus NNS faculty

	NS		NNS		
Rating scales	*M*	*SD*	*M*	*SD*	*t*
Holistic (100)	82.25	(8.02)	72.50	(5.58)	2.09
TWE	4.13	(.85)	3.35	(.47)	2.21*
ESL Composition Profile (Analytic 1)	75.48	(6.43)	72.40	(4.79)	.99
2nd analytic scale	62.75	(18.12)	55.10	(8.85)	1.09

Table 3. Rating outcomes of MA faculty versus PhD faculty

	MA		PhD		
Rating scales	*M*	*SD*	*M*	*SD*	*t*
Holistic (100)	70.73	(8.50)	75.91	(6.53)	1.24
TWE	3.17	(.29)	3.68	(.72)	1.19
ESL Composition Profile (Analytic 1)	73.67	(3.22)	73.17	(5.80)	-.14
2nd analytic scale	56.67	(4.16)	57.45	(13.49)	.10

Table 4. Descriptive statistics for academic disciplines

	Academic Discipline					
	Linguistics		Literature		TESOL	
Rating Scales	*M*	*SD*	*M*	*SD*	*M*	*SD*
Holistic (100)	77.6	8.44	73.25	9.07	73.00	3.46
TWE	3.9	.89	3.38	.48	3.4	.55
ESL Composition Profile	75.18	4.52	74.5	6.25	70.4	4.83
2nd analytic scale	63.8	11.76	54.25	13.23	53.20	10.57

Table 5. One-way ANOVA for academic discipline

Rating scales	Source	SS	df	MS	F	p
Holistic 100	Between	64.91	2	32.45	0.62	>.05
	Within	579.95	11	52.73		
TWE	Between	.84	2	.42	.91	>.05
	Within	5.09	11	.46		
Analytic 1	Between	65.48	2	32.74	1.23	>.05
	Within	292.05	11	26.55		
Analytic 2	Between	258.96	2	129.48	.82	>.05
	Within	1742.75	11	158.43		

Table 6 Rating outcomes of different rating scales

Rating Scales	L	H	Mean	SD
Holistic (100)	62	88	74.7	7.04
TWE(transformed)	50	83	59.5	11.3
Analytic 1	64	81	73.3	5.24
Analytic 2	37	79	57.3	11.95

$p < .05$. Friedman test $\chi_2 = 30.500$, $df = 3$ $p < .000$

Table 7. Relations between Holistic Scores and Components of Analytic Scores

	Components of Analytic Scores									
Holistic	A1cont.	A1gr	A1org	A1voc	A1mec	A2cont.	A2gr	A2org	A2voc	A2mec
(100)	.537*	.046	.537*	.387	.446	.396	.209	.616*	.196	.677*
TWE	.588*	.290	.600*	.335	.471	.593*	.086	.550*	.141	.412

Table 8. Categories of comments made by raters

Categories of comments made by raters	
A. Introduction technique	I. Transition
B. Organization	J. Pronoun
C. Content	K. Relevance
D. Details & development	L. Conclusion technique
E. Grammar	M. Personal grading criteria
F. Vocabulary variety	N. Mechanics
G. Word choice	O. Other
H. Meaning	

Table 9. Rater comments

	NS				NNS									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	Linguistics		TESOL	Lit	Linguistics			Literature			TESOL			
Total number of comments	35	23	25	21	46	16	20	24	36	11	48	28	13	31
Number of factors commented on	9	7	13	14	14	6	7	11	14	6	12	13	5	11
Language use	25	21	18	6	22	10	16	17	17	9	21	15	11	17
Contents	4	2	2	8	9	2	2	4	4	0	14	8	1	8
Organization	4	0	2	3	8	4	2	3	4	2	8	4	1	7
Holistic(100)	88	85	78	70	70	75	70	82	62	79	72	75	70	70
TWE	5	4.5	3	4	4	3	3	4	3.5	3	3	3	4	3
Analytic 1	79	80	66	77	76	71	70	81	76	75	68	71	64	72
Analytic 2	79	70	37	65	42	64	58	68	60	52	62	42	54	43

BEST COPY AVAILABLE